数理解析研究所講究録
第2065巻 2018年 140-146

ネットワーク障害発生の定量的評価への誕生日重複問題の非線形近似理論的方法

東京理科大学理工学部情報科学科　児玉賢史
東京理科大学理工学部情報科学科　Tong Yao
東京理科大学理工学部情報科学科　明石重男

１．序章

『６０人のクラスで、同じ誕生日を持つ生徒が２人以上存在する確率は、９９パーセント以上である』という話を聴いたことがある人は多いと思われる。しかし、この話を厳密に解釈するためには、『９９パーセント以上の確率』という言葉の意味をもう少し正確に理解しておく必要がある。この言葉は、『６０人のグループを１００組用意したとき、同じ誕生日を持つ人たちがいるグループが約９９組も存在する』ということである。一般に１年は３６５日あるため、感覚としては１００人くらい集めないと同じ誕生日を持つ人たちがいないように思われることから、この結果は意外に感じてしまう。ちなみに『２３人のクラスで、同じ誕生日を持つ生徒が２人以上存在する確率は、約５０パーセントである』という結果も有名な話である。

フグという魚が内臓に毒を持っていることは良く知られているが、与えるえさによってある程度まで減らすことができるらしい。話の真偽は別として、このような仮定で進めた場合、『フグの持つ毒を５パーセントまで減らすことが可能になった』という内容は、２通りの意味に解釈することが可能である。つまり、『任意に選び出したフグに含まれる毒を調べてみたところ、通常のフグの毒の５パーセント分しか含まれていなかった』という解釈と、『２０匹のフグを任意に選び出して調べたとき、１９匹のフグは無毒であったが１匹だけが通常のフグと同じ含有量の毒を持っていた』という解釈が考えられる。同じ５パーセントでも、この２種類の解釈には、現実的に大きな相違点が存在する。例えば、『フグに含まれる毒が５パーセントに減らされた場合に、食用にしても安全である』ことを仮定すると、前者の解釈の場合、全てのフグを食用にしても安全であることになる。しかしながら、後者の解釈の場合では、２０匹のフグのうち１９匹は食用にしても安全であるが、残りの１匹については、以前と同じ危険性が残ることになる。このように、代表値としての確率論的数値の持つ意味を正確に解釈することは、非常に重要であり、同じようなことが『大都市で排出され光化学スモッグの原因となっている自動車の排気ガスの５パーセント削減』などという話についても適用できることが知られている。

２．表計算ソフトを用いた誕生日重複問題への解答

『６０人のクラスで誕生日の重複が起きない確率』をＰとおく。このとき『６０人のクラスで同じ誕生日を持つ生徒が２人以上存在する確率』は、１－Ｐで与えられる。従って、正確なＰの値は、例えば表計算ソフト等を用いることで、前述の問題の結果を得ることができる。具体的には、

$$P = (1-0/365)*(1-1/365)*\cdot\cdot\cdot(1-59/365) = 0.005893$$

を求めればよく、これは、一般的な表計算ソフトを利用すれば容易に出力可能である。以下の表１.に順次計算した結果を示す。

表１．表計算ソフトを用いて演算した場合

グループ構成人数	メンバー全ての誕生日が異なる確率	誕生日重複者の存在確率
1	1	0
2	0.997260274	0.002739726
3	0.991795834	0.008204166
4	0.983644088	0.016355912
5	0.972864426	0.027135574
6	0.959537516	0.040462484
7	0.943764297	0.056235703
8	0.925664708	0.074335292
9	0.905376166	0.094623834
10	0.883051822	0.116948178
11	0.858858622	0.141141378
12	0.832975211	0.167024789
13	0.805589725	0.194410275
14	0.776897488	0.223102512
15	0.74709868	0.25290132
16	0.716395995	0.283604005
17	0.684992335	0.315007665
18	0.653088582	0.346911418
19	0.620881474	0.379118526
20	0.588561616	0.411438384
21	0.556311665	0.443688335
22	0.524304692	0.475695308
23	0.492702766	0.507297234
・・・	・・・	・・・
59	0.007010552	0.992989448
60	0.005877339	0.994122661

以上で、『１グループ２３人で約５０パーセント』であること、さらに『６０人で約９９パーセント』であることが確認することができた。次の章では、表計算ソフトを用いない計算方法を提示する。

３．区分求積近似法を用いた誕生日重複問題への解答

前節でも述べた『誕生日重複問題』は、『常識では意外に思えるが、真実であるような事柄に対して、数学的に正当化できる説明を与える確率論』の立場では、数値シミュレーション的役割を果たしている。しかし表計算ソフトによる演算は、あくまで数値実験的傾

向が強く、今回のような場合は、単純な『確認的役割』になってしまう。以下では、『P の近似値』を、区分求積法を用いて求めることを考える。最初に、

$$P = \Pi_{k=0}^{59} (1-k/365)$$

に対して log 演算を適用した後、積演算を和演算に置き換え、365 で割ることにより、以下のように刻み幅 1／365 の区分求積近似計算式が得られる。具体的には、

$$\begin{aligned}(1/365)*\log P &= (1/365)\Sigma_{k=0}^{59} \log(1-k/365)\\ &\fallingdotseq \int_{[0,\,60/365]} \log(1-x)\,dx\\ &= [(x-1)\log(1-x) - x]_{[0,\,60/365]}\\ &= (-305/365)\log(305/365) - 60/365\end{aligned}$$

という結果が得られることになる。上の式における２番目の等号が、区分求積を定積分によって近似していることに注意して、

$$\begin{aligned}&305/365 = 0.835616\\ &\log(305/365) = -0.179590\\ &-305\log(305/365) = 54.7736\end{aligned}$$

という結果を代入すると、

$$P\text{の近似値} = \exp[-60 - 305\log(305/365)] = 0.005373$$

が得られ、この値は、前節で表計算ソフトを用いて計算した値に近いことが確認される。このような近似計算が可能である理由として、次の２点を挙げることができる。

(1) １／３６５ ずつ減少する階段関数に対して、『非線形関数である対数関数』を用いた近似を行ったこと。
(2) 刻み幅が １／３６５という小さな値であることに注目して、『階段関数の面積計算』を、『階段関数を近似する対数関数の定積分計算』に帰着させたこと。（通常は面積を階段関数で近似することが行われる。）

今回の例にもあてはまることだが、一般に不連続的に変化する関数を連続関数で近似する場合は、線形関数で近似する場合よりも非線形関数で近似する場合の方が有効性が高い。そのような具体例として、ファン・デル・ワールスの方程式と呼ばれる『気体の状態方程式』が存在する。一般に、理想気体では気体の体積と絶対温度とは正比例関係にあるが、実在気体では、分子間力などの影響を無視できず、高温になるほど、分子運動が激しくなるため、必ずしも正比例や反比例関係にはならないことが知られている。

4．ネットワーク孤立地域発生に関する数学的評価

計算機を用いて実際に数値を計算する場合、通常の数学では考えられないようなことが起こり得る。例えば、『区分求積分を用いた定積分の近似計算』においては、『刻み幅を小さく取れば取るほど近似精度が向上する』と思われがちだが、そうならない場合があり、また、『非常に大きな数を用いた計算』においても、『一の位や十の位などの値が切り捨てられてしまう』という桁落ち現象が発生するといった場合がある。このような事実は、計算機を用いた数値計算の世界では、『非常に大きな数値』や『非常に小さな数値』が表れないような演算処理を考えなくてはならないということを意味している。ここで順列と組み合わせに関して、『n 個の中から、重複を認めないで r 個選び出す場合の数』に対する『m個の中から、重複を認めないで r 個選び出す場合の数』の比率を求めることを考えた場合、具体的には、

$$mCr／nCr = [m(m-1)\cdot\cdot\cdot(m-(r-1))]／[n(n-1)\cdot\cdot\cdot(n-(r-1))]$$

を計算することになる。そこで、$r \leqq m \leqq n$ と仮定して、$I(n,m,r)$ を次の式で定義する。

$$I(n,m,r) = (1／n)\Sigma_{k=0}^{r-1} \log(m／n-k／n)$$

ここで、先程と同様の区分求積法による近似を用いると、n が十分大きいとき、

$$I(n,m,r) \fallingdotseq \int_{[0,\,r／n]} \log(m／n-x)\,dx$$

という式が得られる。数学的に厳密に述べるならば、s および t を $s \leqq t$ を満たす正定数として、$s = r／n$ および $t = m／n$ という関係が成立するという仮定の下で n を十分大きい自然数とした場合、

$$\begin{aligned}\lim_{n\to\infty} I(n, nt, ns) &= \int_{[0,\,s]} \log(t-x)\,dx \\ &= (s-t)\log(t-s) + t\log t - s\end{aligned}$$

が成立することを意味している。従って、

$$(1／n)\log(mCr／nCr) \fallingdotseq I(n,m,r) - I(n,n,r)$$

という関係が得られるので、

$$\lim_{n\to\infty} (1／n)\log(ntCns／nCns) = (s-t)\log(t-s) + t\log t - (s-1)\log(1-s)$$

が導かれる。この式は、例えば「『１から９０までの数字を書いた９０枚のカードの中から１０枚を選ぶ場合の数』に対する『１から１００までの数字を書いた１００枚のカードの中から１０枚を選ぶ場合の数』の比率が求まれば、『１から９００までの数字を書いた９００枚のカードの中から１００枚を選ぶ場合の数』に対する『１から１０００までの数字を書いた１０００枚のカードの中から１００枚を選ぶ場合の数』の比率も、近似的に計算可能である」ということを示している。

ここで、上式の左辺に関する具体的計算例を以下に示す。表２.では、$s = 0.1$ および $t = 0.9$ とした場合の計算結果をまとめたものである。n が大きくなるにつれて、ntCns／nCns の値が、s と t のみで決定される値に収束することが確認可能である。

表２．近似解の演算結果

n の値	100	200	300	400
(1／n) log(ntCns／nCns)	-0.11072	-0.11103	-0.11113	-0.11119

一方、上記の右辺に関する極限値は、

$$(s-t)\log(t-s) + t\log t - (s-1)\log(1-s) \fallingdotseq -0.1113$$

となるので、数値実験からも、この極限式が成立していることが示された。なお、この方法は、一般化された組み合わせにも拡張することができる。例えば、「４種類の商品 P1,P2,・・・、Q1,Q2,・・・、R1,R2,・・・、S1,S2,・・・を、それぞれ p 個、q 個、r 個、s 個選んで組み合わせる場合の数」は、[(p+q+r+s)!]／[p!q!r!s!] 通りであり、

$$(p+q+r+s)C(q+r+s)\cdot(q+r+s)C(r+s)\cdot(r+s)Cs=(p+q+r+s)Cp\cdot(q+r+s)Cq\cdot(r+s)Cr$$

と一致するため、上式を組み合わせることで、p+q+r+s が十分大きい場合の組み合わせ総数計算にも応用することが可能である。

５．ネットワークの頑健性評価へ簡単な応用

下の図１.は、パソコンとルーターを用いて相互通信するための一例を表している。ルーターには、他のルーターを接続するための「ルーター用インターフェイス」が２個と、パソコンを接続するための「パソコン用インターフェイス」という機材が１個備え付けられていて、パソコンも、ルーターなどのネットワーク機器と接続するための「インターフェイス」を１個備えているものとする。当然であるが、ルーターに備え付けられたインターフェイスが故障すると、そのインターフェイスと繋がっている反対側のインターフェイスを持つルーターやパソコンとは通信ができなくなってしまう。このような状況下では、『１台のルーターに含まれる２個のルーター用インターフェイスが同時に故障する状況』は、『そのルーターに接続しているパソコンが自分以外のいずれのパソコンとも通信不可能となる状況』を引き起こす。下の図１.では、３台のルーターと３台のパソコンが正３

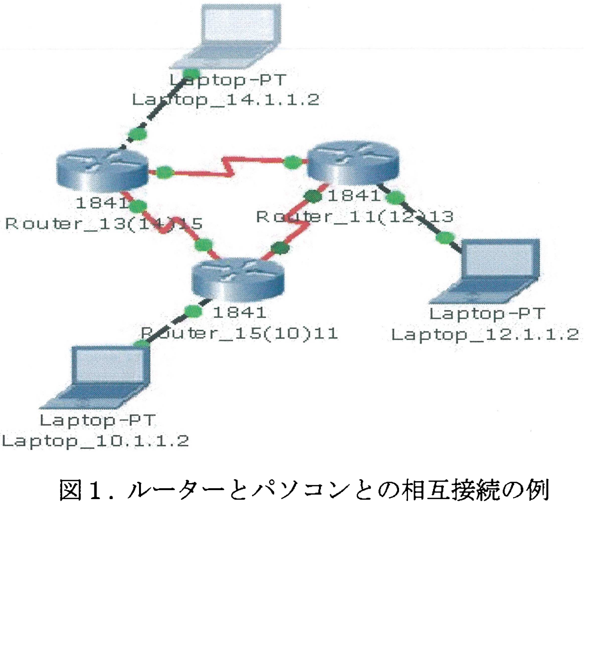

図１．ルーターとパソコンとの相互接続の例

角形のリング状に配置されており、図中の丸印の部分がインターフェイスを表している。

ここで、上述の設定を拡張する形で、リング状に設置された３６５台のルーターのそれぞれに１台のパソコンが接続されている状況下を考える。このとき『２３個のインターフェイスが故障した』という条件の下で、『１台のルーター内部で、同時に２個の接続インターフェイスが故障する条件付確率』はどのようにして計算できるかという『インターフェイス故障に伴うネットワーク接続環境の頑健性問題』を考えることにする。ちなみに、『同時に２個のインターフェイスが故障する状況』というのは、『そのルーターに接続されているパソコンが、両サイドのパソコンと通信ができなくなってしまう状況』つまり『他のパソコンと通信不可能になる状況』を意味しているが、この問題は、先に述べた『誕生日重複問題』に帰着することができる。具体的には、

(1) ３６５台のルーター ＝ １年３６５日

(2) ２３個の故障したインターフェイス ＝ ２３人のグループ

(3) ２個のインターフェイスが同時に故障しているルーターの存在

＝誕生日が共通している２人の存在

という対応付けにより、リングタイプのネットワークトポロジーにおけるネットワーク孤立地域の発生状況は、誕生日重複問題へ帰着できることがわかる。更に次の図２.では、冗長経路を想定して、左右方向に大動脈としての２本のルートを有するネットワークを考えているが、楕円形で囲まれた部分の４個のルーターのうち、１台のルーターにおけるインターフェイスが２個同時に故障した場合、別のルーターに、膨大な負荷がかかり、結果として、データ転送の遅延を引き起こすことが考えられる。

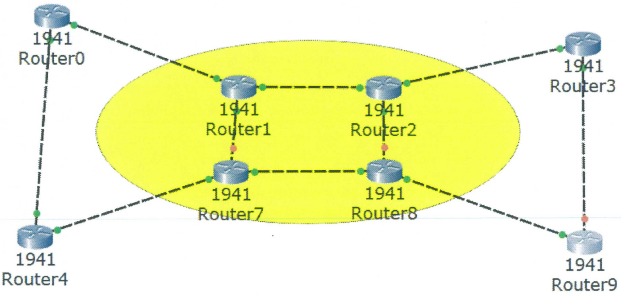

図２．ルーター同士の相互接続の例

このような冗長経路を設定したネットワークでは、今回の問題は、ネットワーク輻輳を引き起こす確率を定量的に評価することに応用できると考えられる。

参考文献

[1] 明石 重男, 児玉 賢史, 桜井 良子, お誕生日重複問題への非線形近似法の応用, 理大科学フォーラム, 2011(12), 44-48.

[2] 株式会社ソキウス・ジャパン編著, 徹底攻略 Cisco CCNP ROUTE 教科書, インプレスジャパン, 2012 年.

[3] 株式会社ソキウス・ジャパン編著, 徹底攻略 Cisco CCNP SWITCH 教科書, インプレスジャパン, 2012 年.